Gartner Licensed for Distribution

Magic Quadrant for Augmented Data Quality Solutions

11 February 2026 - ID G00832289 - 62 min read

By Sue Waite, Divya Radhakrishnan, and 1 more

Augmented data quality solutions detect and fix errors, remove duplicates, standardize formats, and validate data so it can be trusted for business operations, reporting and decision making. The research helps D&A leaders understand these AI-enhanced solutions to make better purchasing decisions.

Strategic Planning Assumptions

- By 2027, 70% of organizations will adopt modern data quality solutions to better support their AI adoption and digital business initiatives.
- By 2027, 80% of mainstream data quality vendors will leverage large language models (LLMs) and natural language processing (NLP) to enable interactive user inference to improve user productivity and tool efficiency.
- By 2027, 70% of organizations seeking data and analytics (D&A) governance tools will prioritize automation, choosing solutions that minimize manual intervention and maximize efficiency.
- By 2030, 80% of organizations will adopt a unified approach for structured and unstructured governance, driven by AI.

Market Definition/Description

*This document was revised on 12 February 2026. The document you are viewing is the corrected version. For more information, see the **Corrections** page on gartner.com.*

Gartner defines augmented data quality (ADQ) solutions as a set of capabilities that deliver advanced features to streamline the identification of quality issues, offer context-aware suggestions for corrective actions, and automate key data-quality processes to ensure cleaner, more reliable data. These purpose-built data-quality solutions support profiling and monitoring, rule discovery and creation, active metadata use, data transformation, data remediation, matching, linking and merging, and role-based usability. The solutions have AI-assistant-enabled features that enhance user experience.

Packaged ADQ solutions help implement and support the practice of data quality assurance, mostly embedded as part of a broader data and analytics (D&A) strategy. Typical use-case scenarios include:

- **Analytics and AI readiness:** Data-quality capabilities supporting the preparation and ongoing monitoring of structured, semistructured and unstructured data for operational analytics, performance management, sentiment analysis, improving the quality of data used for training AI models or algorithms, and actual data feeds to production.
- **Data engineering:** Data-quality capabilities supporting various key data processing in the context of data engineering initiatives, which include general data integration or data migration scenarios.
- **D&A governance:** Data-quality capabilities supporting the data governance initiative and its associated key roles (such as chief data and analytics officers [CDAOs] and data stewards) with a focus on increasing the value of data assets while managing risks and compliance.
- **Master data management** (MDM)**:** Data-quality capabilities supporting various key master data domains in the context of MDM initiatives and the deployment of custom or packaged MDM solutions.
- **Operational/transactional data quality:** Data-quality capabilities support control over the quality of data created by, maintained by, and housed in operational/transactional applications, including Internet of Things (IoT) systems.

Mandatory Features

The mandatory features for this market include:

- **Connectivity:** The ability to access and apply data quality across a wide range of data sources, including internal/external, structured/semistructured/unstructured, at-rest/streaming, on-premises/cloud, and relational/nonrelational data sources.
- **Profiling and monitoring/detection:** The statistical analysis of diverse datasets (ranging from structured to unstructured data and from on-premises to cloud) to give business users insights into the quality of data and to enable them to identify data-quality issues. Profiling results drive the ongoing monitoring for data quality issues based on preconfigured or custom-built monitoring rules (or adaptive rules) and alert violations. Active and passive metadata inform data type recognition and support automatic detection of outliers, anomalies, patterns, and drifts.
- **Rule discovery, creation, and management:** The ability to discover, recommend, design, deploy, and manage business rules for specific data values throughout the life cycle of these rules. The rules can be called within the solution or by third-party applications for data validation or transformation purposes, and the rules may be executed in batch or real-time mode. Augmented solutions support the creation of data-quality rules by converting natural language descriptions of data-quality requirements into executable code. Solutions use active metadata and data analysis to infer relationships and dependencies and to automatically suggest new data-quality rules.
- **Alerts, notifications, and visualization:** The interactive analytical workflow and visual output of statistical analysis help business and IT users identify, understand, and monitor data-quality issues and discover patterns and trends over time (e.g., through reports, scorecards, dashboards, and mobile devices). Based on the anomalies detected, users receive recommendations about new alerts to add. Solutions should actively learn which issues are irrelevant based on user behavior and feedback, then refine generated notifications accordingly.
- **Data transformations (parsing, cleansing, and standardizing data):** The decomposition and formatting of diverse datasets based on government, industry, or local standards, business rules, knowledge bases, metadata, and ML. This feature also involves modifying

data values to comply with domain restrictions, integrity constraints, or other business rules. Augmented solutions use a combination of supervised and semisupervised AI and ML models or LLMs to parse, standardize, and cleanse data.

- **Matching, linking, and merging:** Matching, linking, and merging related data entries within or across diverse datasets using a variety of traditional and new approaches, such as rules, algorithms, metadata, AI, and ML. Augmented solutions use AI/ML/LLMs to suggest potential matches automatically and can tune the results based on user feedback. For merging tasks, consolidation rules for merging data are automatically suggested and refined based on user feedback. Minimal user involvement is required in selecting algorithms, constructing specific match and consolidation rules, and configuring and tuning match parameters.

- **Usability, workflow and issue resolution:** The solution's suitability to engage and support both technical and nontechnical roles required in a data-quality initiative. A workflow includes processes and a user interface to manage data-quality issue resolution through the stewardship workflow, and to enable business users to easily identify, quarantine, assign, escalate, and resolve data-quality issues as facilitated by collaboration, pervasive monitoring, and case management. Augmented solutions initiate and assign data-quality issues by leveraging and activating business, technical, and operational metadata.

- **Active metadata and lineage support:** The ability to collect, discover, or import active and passive metadata from third-party tools and to build or import lineage to perform rapid root cause analysis of data-quality issues and impact analysis of remediation. This feature includes a metrics view based on critical data elements.

Common Features

The common features for the stand-alone or unified data-quality platform market include:

- **Data validation and enrichment:** Augmented solutions support integration with third-party AI models, such as LLMs, to validate or enrich datasets. They can also integrate externally sourced data to improve accuracy and completeness, or add value.

- **Address validation/geocoding:** Capabilities supporting location-related data standardization and cleansing, as well as completing partial data in real-time or in batch processes.

- **Multidomain support:** The ability to address multiple data subject areas (such as various master data domains and vertical industry domains) and depth of packaged support (e.g., prebuilt data-quality rules) for these subject areas. For augmented solutions, multidomain support is coupled with the ability to automatically recommend or deploy prepackaged content based on the explicit or inferred semantics of the data being profiled.
- **AI-assistant-enabled interactions:** These AI-powered conversational agents facilitate self-service data-quality activities (e.g., queries, creating and refining data-quality rules, configuring data remediations, and implementing workflow actions).
- **Unstructured data support:** The ability to analyze unstructured or semistructured data to highlight data-quality issues based on semantic analysis and business validation logic, and generate context-specific metadata. AI/ML capabilities are leveraged to validate the accuracy, completeness, and consistency of unstructured data by assessing data quality based on the availability and completeness of metadata by applying specific validation logic. Unstructured data support also aids in data preparation using NLP techniques to extract information based on entity recognition and sentiment analysis, parsing techniques to extract data and prevent noise, and graph technologies to identify relationships across the extracted entities.
- **Deployment environment, architecture:** Deployment styles include hardware and operating system options, configuration of data-quality operations and processes, and interoperability with third-party tools.
- **Integration with third-party tools:** Augmented solutions can integrate observability and monitoring metrics with other tools to improve holistic notification, analysis, and orchestration of issues with data flows.

Magic Quadrant

Figure 1: Magic Quadrant for Augmented Data Quality Solutions

Gartner

Vendor Strengths and Cautions

Ab Initio Software

Ab Initio Software is a Leader in this Magic Quadrant, headquartered in Lexington, Massachusetts, U.S. Its data quality product is Ab Initio Data Quality Environment (DQE), which is part of the Ab Initio Data Platform. Ab Initio Data Platform covers data integration, data quality, metadata management, data governance, and AI Central for agentic AI orchestration and autoremediation.

Ab Initio's DQE product has about 2,000 customers across sectors, including financial services and insurance, telecom and healthcare. Its operations are geographically diverse and the company primarily targets mid- or large-sized enterprises with significant on-premises data landscapes. Future focus is on AI and autonomy with agentic workflows for autonomous governance, predictive DQ and anomaly prevention, with guardian agents to monitor policy compliance.

Strengths

- **AI and agentic innovation:** Ab Initio introduced "AI Central," a framework for agentic workflows. AI agents autonomously detect, trace and remediate data quality defects without human intervention. As new datasets are registered, the platform's "Day Zero Data Quality" automates creation of statistical profiles and business asset definitions. DQE provides AI-assisted support for processing all types of semi- or unstructured content.

- **Comprehensive solution:** Data quality is an integrated capability within the Ab Initio Data Platform. The platform uses a specification style framework to separate logical control definitions from physical implementations. Users may recompile and retarget rules for different environments without reworking business definitions. Its architecture is portable across hyperscalers (AWS, Azure, GCP) and supports containerization. Client reviews found the solution powerful for large data volumes and heavy processing scenarios. Ab Initio was selected "Customer's Choice" in Gartner's Voice of the Customer research.

- **Financial Stability and Operational Viability:** Ab Initio reports revenue growth every year since 1997 and operates with no long-term debt or external stakeholders. The company derives 100% of its revenue from its data management platform. Multiyear agreements provide over 70% of its revenue. Ab Initio reports that around 85% of pilots convert to sales.

Cautions

- **Focus on technical over business buyers:** Ab Initio is refocusing from technical buyers to business buyers. While it is redesigning its user experience, it has lagged the market in elevating capabilities from low-level developer interfaces to those meaningful for nontechnical users.

- **Lengthy sales and implementation cycles:** The sales strategy relies heavily on “proof of value” pilots and “test drives” rather than quick-start deployment, which dictates a longer timeline to finalize enterprise agreements.
- **Learning curve and self-service resources:** Gartner Peer Reviews indicated that Ab Initio has a steep learning curve. Companies also stated that limited public documentation, public communities/forums, tutorials or other sources make it more challenging to find self-guided resources.

Acceldata

Acceldata is a Niche Player in this Magic Quadrant, headquartered in Campbell, California, U.S. Its data quality product is part of Acceldata Agentic Data Management (ADM), built on Acceldata Data Observability Cloud (ADOC), which provides data observability, validation, reconciliation, lineage and anomaly detection.

Acceldata has around 60 customers for its data quality products. Its operations are primarily in North America, with a few in EMEA and APAC, and primarily targets large enterprises and enterprises managing complex data pipelines. Customers cross sectors, including financial services, life sciences and consumer goods. Future focus is on autonomous, “human-out-of-the-loop” data operations, with self-healing DQ workflows.

Acceldata did not respond to requests to review the draft contents of this document.

Strengths

- **AI-assistance for data quality and data reliability:** AI-driven agents generate data quality policies and automate root cause analysis. ADM provides root cause analysis (RCA) through the solution’s integration with xLake Reasoning Engine. The xLake Reasoning Engine helps ADM support data observability, data validation and reasoning activities, via its AI agents. ADM supports large-scale data validations on petabytes of data/billions of rows using full datasets versus via sampling.
- **Usability and workflow:** ADM automatically creates remediation workflows to support correction of quality issues. If data is missing, ADM launches enrichment workflows that complete or infer data from trusted sources. Data stewards may describe issues in plain language and the AI assistant provides diagnostics, causes and solutions.

- **Metadata use, lineage, and anomaly detection:** Acceldata uses a knowledge graph approach to link technical metadata with business context, which further enhances lineage visibility. Acceldata uses multivariate machine learning (techniques to analyze and model datasets with multiple variables simultaneously) to identify most quality issues with zero manual setup required.

Cautions

- **Market Presence and mind share:** Acceldata is a relatively small data quality vendor with about 60 customers, the majority based in North America. Acceldata currently has 100% direct sales and does not use VARs or distributed partnerships, which limits the vendor's reach and presence.
- **Limited data transformation functions:** Acceldata's approach for transformations relies on coding via custom SQL. Out-of-the-box prebuilt advanced transformation features are not available today. Matching features are defined via user-defined codes versus through the UX/UI and via intelligent configurations. Acceldata does not support entity resolution or householding capabilities, relying on its partnership with Reltio for rich match and merge support.
- **Unstructured content support:** Acceldata currently supports only semistructured content: XML, JSON, Office documents. All others are roadmap items: binary (images, audio, video, GIS data) textual (PDF, invoices, claims data, genomics data, CMS, research papers, Wiki, blogs, legal documents).

Anomalo

Anomalo is a Niche Player in this Magic Quadrant, headquartered in Palo Alto, California, U.S. Its data quality product is also named Anomalo. The Anomalo platform provides automated data quality monitoring and anomaly detection via unsupervised machine learning to identify and resolve issues across structured and unstructured data.

The vendor has over 70 customers located mostly in North America, with some in EMEA and APAC. It primarily serves the financial services and insurance, retail and consumer goods, and communications/media/entertainment sectors. Anomalo targets large enterprises and Fortune 500 companies with mature data strategies. Future focus prioritizes AI intelligence

and expanding access to that intelligence with new enhancements for unstructured content and AIDA, Anomalo's agentic interface.

Strengths

- **Expanded unstructured data support:** Anomalo enriched its support for unstructured content and files to help users detect missing, delayed or malformed files before they impact downstream AI/RAG processing or regulatory requirements, including redacting sensitive content identified in files. Unstructured content is ingested as collections, with 15-plus automated quality checks, then Anomalo automatically applies no-code extraction of fields, metadata and insights. Workflows support the transformation of files into clean, structured outputs for analytics, compliance and AI.
- **Data profiling and monitoring:** Anomalo uses unsupervised ML to learn data behavior and patterns, enabling automatic monitoring versus requiring users to write rules or set thresholds. Anomalo detects significant statistical shifts, drifts or outliers that users typically would not look for.
- **Customer responsiveness:** Anomalo releases weekly solution updates, and 30% of all enhancements are directly based on clients' feedback and suggestions. Its clients highlight Anomalo's responsiveness as a benefit of working with this vendor.

Cautions

- **Data lakehouse dependency for lineage:** Anomalo renders lineage only for tables it monitors within Databricks, Snowflake and BigQuery. Anomalo also supports manual integration of other lineage details via API; however, it lacks the native, cross-platform lineage support offered by comprehensive data quality solutions.
- **Limited data transformation support:** Anomalo does not provide point-of-entry data quality standardization, cleansing or matching functions that are executed live within the native business application's environment (CRM entries, ERP entries). Anomalo does not offer address validation and enrichment, nor geocode enrichment services. Anomalo's transformation capabilities are not designed for interactive, business-user-driven data cleansing.

- **Visualizations:** Gartner's Peer Insights reveal that some companies feel that Anomalo's out-of-the-box visualizations are overly complicated and that it requires time for business users to learn how to read them.

Ataccama

Ataccama is a Leader in this Magic Quadrant, headquartered in Boston, Massachusetts, U.S. Its data quality product is Ataccama ONE Data Quality Suite, which is part of the Ataccama ONE platform. The platform provides integrated data quality and governance capabilities, as well as metadata management, data observability and master data management.

Ataccama has about 580 customers for its data quality products, with most located in North America, followed by EMEA. Its clients are primarily in the financial services, manufacturing and insurance sectors. Future focus is to extend its AI agent's capabilities, expand data observability toward AI models, and activate unstructured content for AI use via vector stores.

Strengths

- **Integrated agentic capabilities**: Ataccama introduced AI agent features capable of using chain-of-thought reasoning to create plans and execute tasks across the platform. This AI capability supports the automation of rule creation, data documentation and issue remediation, allowing users to interact with data using natural language. The Ataccama ONE platform also includes specific AI-driven features for unstructured data activation and automated data quality gate enforcement in pipelines.

- **Data Trust Index**: Ataccama introduced this function as a multidimensional key performance indicator (KPI) to help users understand the trustworthiness of a dataset. The index provides a calculated score based on data quality, metadata completeness, observability, lineage, governance and data adoption. Future plans include refinement of scores based on industry-specific content packs.

- **Transparent, capacity-based licensing:** Ataccama offers a simplified subscription licensing model based on processing capacity and number of named users across product suites. This includes all features for the selected product suite, such as the Data Quality & Governance Suite. It includes Ataccama ONE platformwide capabilities, such as

ONE AI Agent, integrated across suites. All tiers provide unlimited read-only consumer users and connectors without additional costs.

Cautions

- **Implementation support:** Gartner's Peer Insights reflect that Ataccama clients have had very limited choice in implementation partners. Ataccama very recently began expanding its global system integrators (GSI) network, providing more flexibility and choice for clients. Companies that prefer to work with a specific GSI should confirm the depth of that partner's experience with the Ataccama ONE platform.
- **Primary focus on large enterprise:** Ataccama's ideal customer profile is focused on larger enterprises, particularly in regulated industries like banking and insurance. This focus may impact availability and accessibility for midmarket organizations' needs.
- **Limited customer base growth YoY:** Compared with other vendors in this research, Ataccama experienced relatively lower customer growth. New customers were added for current solutions, with some attrition in legacy solution customers.

CluedIn

CluedIn is a Niche Player in this Magic Quadrant, headquartered in Copenhagen, Denmark. Its data quality product is named CluedIn Data Quality, which is one of the offered modules as part of its unified CluedIn Agentic Data Management platform. CluedIn's graph-based platform automates data profiling, cleansing and deduplication for both business and technical users.

Its data quality product currently has over 110 customers, which are mostly in EMEA, with some in North America and APAC. CluedIn targets large and medium-sized enterprises. Its customers are primarily in the financial services, insurance and manufacturing sectors. Future focus is to extend AI agent features to update source applications, expand its multitenant SaaS footprint to AWS and GCP, and launch an open-source initiative.

Strengths

- **AI assistance:** CluedIn includes an AI agent-based workforce approach to extend data quality capabilities through use of LLMs to perform profiling, data quality rules generation, data quality cleansing and matching, and support workflows. CluedIn

provides features to configure and deploy custom AI agents for DQ tasks. CluedIn offers a free public SaaS option, up to 15,000 records, to trial its solution with use of all features, including access to its AI agents and CoPilot.

- **User experience:** Gartner Peer Insights reviewers praise CluedIn's ease of use, intuitive user experience that it is fast to deploy with out-of-the-box connectivity (zero-upfront modeling support) and ability to ingest data without a predefined schema. Its knowledge graph structure builds relationships dynamically. CluedIn provides detailed costs for DQ agent processing. Clients reported that advanced features require deep understanding to use well.
- **Strategic partnership with Microsoft:** CluedIn continues its strong relationship with Microsoft. CluedIn's offerings are presented as native solutions for master data management and data quality within Microsoft Azure.

Cautions

- **Ecosystem dependency/limited direct sales:** A significant portion of CluedIn's sales pipeline, 54%, is directly derived from the Microsoft Partner Network. CluedIn's commercial foundation remains concentrated on the Azure ecosystem. CluedIn's direct sales force is relatively small, about 10 people.
- **Product strategy:** CluedIn's primary focus remains master data management. Its solution features are primarily focused on data quality activities for master-data-related entities and attributes versus transactional/operational data. Companies that desire an end-to-end data quality solution may prioritize larger, more established and comprehensive platforms.
- **Limited unstructured support:** CluedIn offers only basic support for unstructured/semi structured sources with no current ability to monitor for or track the quality of identified entities. CluedIn indicates this is on its roadmap. CluedIn has no roadmap plans to add support for JSON (IoT data) or binary forms (audio, video, GIS data).

DQLabs

DQLabs is a Visionary in this Magic Quadrant, headquartered in Pasadena, California, U.S. Its data quality product is DQLabs Platform. DQLabs provides a unified augmented data quality

and observability platform that leverages a semantic layer and AI to automate data profiling, anomaly detection and issue remediation.

DQLabs has over 140 active customers for its data quality product, with most located in North America, and some in EMEA and APAC. DQLabs targets large enterprises and strategic accounts via direct sales, as well as midmarket and small/medium enterprises through indirect channels. DQLabs primarily services the banking and securities, technology, and healthcare provider sectors. Future focus is the transition to PRIZM, DQLabs' fourth generation, AI-native platform designed for autonomous "self-driving" execution, with a roadmap focused on AI model observability and cross-cloud interoperability.

Strengths

- **AI for data quality:** DQLabs enhanced its AI augmentation to provide automated data quality functions for rules suggestions and recommendations. Natural language processing may be used to accept/edit those suggestions. AskAI allows users to interrogate data and observability metrics using natural language questions. Profiling for some forms of unstructured/semistructured data (PDFs, documents, XML, JSON) monitors for bias, personal information and accuracy. Potential issues are automatically alerted by DQLabs based on anomalies, drift events and pipeline signals.
- **Support and user experience:** Gartner's Peer Insights reviews highlight DQLabs' excellent support, responsiveness and willingness to collaborate on new feature enhancements. DQLabs' clients also appreciate the solution's ease of use.
- **Diverse persona support:** DQLabs provides an out-of-the-box UI, including NL-enabled interactions, that supports a range of data-quality-related personas, including data stewards, business analysts, business leaders, data engineers and AI/ML teams.

Cautions

- **Emerging partner ecosystem:** Direct sales still account for the majority of DQLabs revenue (74%). DQLabs has relatively recently turned its attention to aggressively expanding its "reseller partner program," to mature their global reach and revenue potential.

- **Regional presence**: DQLabs sales are heavily concentrated in North America (about 70%). Its presence in other regions is significantly lower, which has potential to limit the depth of local support in those regions when compared with other vendors in this research.
- **Limited synthetic data generation:** DQLabs does not currently support native synthetic data generation with enhanced labeling, as needed to support GDPR, HIPAA, and CPRA anonymization for PII and regulated fields (on the roadmap).

Experian

Experian is a Challenger in this Magic Quadrant, headquartered in Dublin, Ireland. Its primary data quality product is Aperture Data Studio, complemented by Experian Governance Studio, Experian Batch, and several data validation and enrichment services such as Experian Address Validation, Experian Email Validation, and Experian Phone Validation. Its data quality solution provides data profiling, standardization, matching and enrichment.

Experian has about 6,100 customers for these products, with most using data validation and contact enrichment services. Its operations are geographically diversified, and its customers are primarily in financial services, retail and public sector. Future plans are to expand its portfolio with prebuilt solution packages for AI governance and automated audits of AI training data, and accelerator packages for compliance regulations, utilities, insurance and housing.

Strengths

- **Proprietary data assets:** Experian's rich history with and access to very large proprietary datasets covering both consumers and businesses continues to be a strong asset for its clients, supporting robust address cleansing and matching. This also allows Experian to support AI training data needs by mirroring real-world context to identify data bias and gaps.
- **AI assistance and match enhancements:** Experian introduced GenAI Actions, which generate data quality rules and workflows from natural language descriptions. Experian also released a new match engine, which now supports both deterministic and

probabilistic matching, hierarchical clusters (how entities relate) and a “cluster review” interface to help resolve ambiguous matches.

- **Business-focused usability:** Experian Aperture supports multiple personas and user types, particularly a “citizen user.” Gartner’s Peer Insights indicated clients are pleased with Aperture’s address cleansing and matching features.

Cautions

- **Evolving cloud maturity:** While Experian is shifting to a multitenant, cloud-native architecture, the company is still in process of moving its large legacy on-premises customer base to the new solutions. Achieving complete architectural modernization across its hosted offerings remains a future roadmap goal. Customers should carefully monitor future releases for alignment to their infrastructure requirements.
- **Limited lineage support.** Experian’s Aperture Data Studio currently provides limited native lineage support. Data lineage is populated by linking databases (and scheduling refreshes) to import schemas, tables and column metadata. Technical lineage is enabled via integration with Solidatus.
- **Inconsistent sales experiences:** Gartner’s Peer Insights revealed that Experian’s clients express a range of experiences during solution evaluation and contract negotiation phases. Gartner client inquiries have also reported challenges with account team interactions, citing a lack of responsiveness.

IBM

IBM is a Leader in this Magic Quadrant, headquartered in Armonk, New York, U.S. Its data quality products are primarily centered around its watsonx family of products, which includes watsonx.data intelligence, watsonx.data integration, and IBM Master Data Management (formerly known as IBM Match 360). In April 2025, IBM repackaged its stand-alone products, InfoSphere QualityStage, DataStage, and Databand into watsonx.data integration; and IBM Knowledge Catalog, Data Product Hub, and Manta Data Lineage, into watsonx.data intelligence.

Gartner estimates more than 3,000 customers for these product lines. IBM primarily targets large enterprises, and its operations are geographically diversified with clients across all

sectors. Future focus is on the development of autonomous data products and the launch of agentic data intelligence tools powered by the Model Context Protocol (MCP) to automate search and compliance tasks.

Strengths

- **AI innovations for data quality:** AI assistants support data quality functions, suggest and recommend actions, and convert natural language statements to technical data quality rules. Data quality insights are integrated with business process visibility to display the impact of quality issues on business outcomes, supporting decision making for different user personas. IBM expanded its approach to data quality for AI via use of data contracts to support establishing data products. "Autonomous data trust agents" will support continuous quality for agentic AI.
- **Unstructured data quality for AI:** IBM provides robust support for all forms of unstructured and semistructured content. Unstructured content is extracted, classified and vectorized. The resulting embeddings are stored in the vector database in support of AI scenarios. IBM provides lineage from the upstream raw documents to the downstream curated AI model inputs.
- **Sales and systems support:** IBM reports strong YoY revenue growth: a triple-digit percentage increase for watsonx.data intelligence, and strong double-digit percentage increase for watsonx.data and watsonx.governance revenue. IBM's watsonx solutions support on-premises, hybrid, multicloud and full SaaS environments.

Cautions

- **Complex portfolio transition:** As IBM transforms from traditional data quality solutions to watsonx-based solutions, existing clients should consider data management requirements and migration challenges. Organizations with siloed legacy systems or poor existing metadata management may find it challenging to leverage IBM's vision of an "active, agent-ready operational fabric" with watsonx for DQ without significant upfront metadata preparations.
- **Constraints for smaller companies:** Client inquiries highlight that IBM solutions often need extensive customization and skilled IT professionals, typically making them more suitable for large enterprises.

- **Self-service product resources:** Gartner's Peer Insights from IBM's clients report difficulty navigating IBM's website to find relevant product details as well as product documentation.

Irion

Irion is a Niche Player in this Magic Quadrant, headquartered in Turin, Italy. Its data quality product is Irion Premium Data Quality & Governance. The platform utilizes a metadata-driven hub-and-spoke architecture to unify data quality, governance, lineage and observability, automating rule management and remediation.

Irion has around 85 customers that are mostly in EMEA, with a small number in North America. Irion primarily services the banking, insurance and utility sectors. Future focus is to enable metadata-aware Data Artificial Intelligent System (DAISY) agents that may autonomously execute data quality actions and the launch of its fully modernized cloud-native architecture.

Strengths

- **User experience:** Irion provides a new graph-based navigation interface that allows users to visually explore data relationships, dependencies and lineage. AI-driven guidance "Know Your Model" (KYM) uses NLP for root cause analysis and impact assessment, simplifying complex data discovery. Per Gartner's Peer Insights, Irion's clients praised its user-friendly UX, its adaptability to different user roles/personas, and the platform's flexibility to rapidly support their organization's data quality activities across many systems.

- **Technology innovation:** Irion's DAISY has been enhanced to use autonomous, metadata-aware agents to detect quality issues, propose data remediations required, and execute corrective actions independently, if desired. Irion offers a low-code rule wizard for use of natural language statements to create and manage rules. Irion provides robust support for almost all forms of unstructured and semistructured content, excluding lidar/radar.

- **Industry-specific regulatory expertise:** Irion provides prebuilt accelerators and validation rules that are tailored to track compliance of complex regulations like BCBS 239 and ECB's RDARR for banking, and Solvency II or IFRS 17 for insurance.

Cautions

- **Geographic presence:** Irion continues to have very limited international presence, with the vast majority of clients based in EMEA. Irion also relies heavily on direct sales, which drive 92% of its revenue. Irion indicates plans to drive its international expansion through strategic partnerships.
- **Performance with large data volumes:** Irion's solution currently relies on MS SQL Server for processing, which can limit data-processing performance with high data volumes, an issue also cited in Gartner's Peer Insights. Irion plans to introduce a new in-memory, columnar analytical engine with full scale-out support early in 2026 to resolve these challenges.
- **Vertical concentration/support:** Irion's revenue is heavily concentrated in banking (51%) and insurance (27%). While this reflects its specialization, it also suggests that support for other industries, via prebuilt content and/or accelerators, is likely less mature and should be carefully evaluated.

Precisely

Precisely is a Challenger in this Magic Quadrant, headquartered in Burlington, Massachusetts, U.S. Its data quality products include Precisely Data Integrity Suite, Precisely Spectrum, Precisely Trillium, and Precisely Data360.

Precisely has 4,970 data quality solution customers. Its operations are geographically diverse, with clients primarily in financial services, insurance and telecommunications sectors. Future focus is on the rollout of Gio AI Assistant, Precisely's conversational AI assistant, and accelerating AI-enabled features through the release of AI agents to automate manual data quality tasks.

Strengths

- **High customer retention:** Precisely reports a 94% customer retention rate for its data quality solutions. Precisely also focuses on expanding its data quality customers through growth within its existing broader customer base of more than 12,000 customers. This accounts for 95% of its data quality revenue, with a YoY growth rate of 8%.

- **Deep location intelligence and data enrichment:** Precisely differentiates its offering through extensive data enrichment capabilities, providing access to over 400 datasets containing more than 9,000 attributes. This is supported by the PreciselyID, a proprietary identifier used to facilitate seamless matching, linking, and enrichment of address and business data across numerous countries and territories.
- **Expertise in challenging environments:** Precisely supports challenging environments such as mainframes and SAP, particularly in point-of-entry data quality cleansing, standardization and matching functions. Precisely's heritage in data integration also supports data integrity assurance for high-volume data in motion or batch across hybrid environments.

Cautions

- **Product strategy and support:** Precisely's portfolio includes distinct legacy "core products" (specifically Trillium, Spectrum, and Data360) that run alongside the Data Integrity Suite. While the vendor supports these products without forced migration, the existence of multiple distinct engines and licensing models (e.g., use-case dependent licensing for core products versus consumption tiers for the suite) creates a complex landscape for customers to navigate compared to single-stack platforms.
- **Lagging AI-assisted data quality support:** Precisely, at the time of this evaluation, provides limited AI-based augmentation for rule creation and business rules management, remediation, reconciliation and exception handling. Companies desiring these types of autonomous quality features should verify current timelines and ensure feature fit for their business requirements.
- **Revenue reliance on existing customers:** A very large portion (95%) of Precisely's data quality revenue comes from its existing customer base. This high reliance suggests that, while Precisely is excellent at upselling current customers, the company will likely face strong competition from AI-native startup data quality vendors when competing for net-new clients.

Qlik

Qlik is a Leader in this Magic Quadrant, headquartered in King of Prussia, Pennsylvania, U.S. Qlik's data quality product is Qlik Talend Cloud, which incorporates features from Talend Data Fabric and Talend Data Catalog. Its metadata-driven platform supports data quality and governance across hybrid environments, using automation and a patented Trust Score to assess data health and AI readiness.

Qlik has over 3,330 active customers for its data quality offering, with operations distributed globally. The highest concentrations are in EMEA and North America. Qlik primarily services the financial services, retail and services, and high-tech industries. Future focus is delivery of contract-governed data products and expansion of FinOps-driven data operations via its Open Lakehouse architecture, following the Upsolver acquisition.

Strengths

- **Market presence and growth:** Qlik demonstrates strong viability with a "Rule of 50" score of 53% (combined ARR growth and profitability) and a 108% net customer retention rate. Gartner's **Market Share Analysis- Data Management Software (Excluding DBMS), Worldwide, 2024** indicates that Qlik's data quality revenue grew by over 32% (36% more than the closest vendor).

- **AI augmentation:** Qlik introduced a GenAI-based rule assistant that suggests rules from schema and profiling. AI suggestions also generate cross-column logic from natural language inputs. As issues are identified, the root cause and which data quality rule(s) triggered the issue identification are included. For metadata collection, dataset, field and product descriptions are automatically suggested for a steward's review, edit and approval.

- **Strong marketing and support:** Qlik's strong global presence across North America, EMEA, and Asia/Pacific and Japan (APJ) are supported by an extensive partner network with more than 8,500 Qlik certified staff, enabling broad implementation and ongoing support options for its customers.

Cautions

- **Migration and brand consolidation complexity:** Qlik is transitioning its large legacy customer base to the Qlik Talend Cloud solution and pricing model. Customers should

assess for feature parity and support across hybrid and multicloud environments while employing careful change management and training.

- **Pricing model change:** Qlik has moved to a usage-based capacity pricing model for Qlik Talend Cloud. This model tracks data volume, job execution and duration factors. While Qlik's new model provides flexibility, companies with unpredictable or high-volume data workloads should use the platform's FinOps tools to carefully monitor usage for unexpected cost impacts.
- **Metadata as a precondition:** Qlik's data quality strategy is heavily metadata-driven, its "Trusted Intelligence Layer" is a precondition for the ability to use its AI-assisted data quality features. Companies with poor existing metadata documentation or siloed legacy systems will likely face a steep learning curve in activating their metadata, to take full advantage of Qlik's AI-augmented DQ features.

Salesforce (Informatica)

Salesforce (Informatica) is a Leader in this Magic Quadrant, headquartered in Redwood City, California, U.S. Its data quality products are Intelligent Data Management Cloud (IDMC), and Informatica Data as a Service for address, email, and phone verification. The IDMC modules referenced in this research include Cloud Data Quality, Cloud Data Governance and Catalog, Master Data Management, and Data Integration.

Informatica has more than 5,000 customers for all product offerings. Its customers are primarily strategic and large enterprises, geographically diverse, and are across all major sectors. Future focus is on AI governance with expansion of its platform to handle agentic workflows, governance of the AI life cycle, and the expansion of data quality controls for unstructured content.

In May 2025, Salesforce entered into an agreement to acquire Informatica. The acquisition closed 18 November 2025.

Strengths

- **Unified, metadata-driven intelligence:** Informatica's IDMC functions as a comprehensive platform powered by its AI engine, CLAIRE, and a unified metadata foundation. Informatica embeds data quality directly into broader workflows like MDM and

governance, ensuring that quality rules defined once can be applied consistently across the platform. The metadata-centric approach supports DQ automation, with CLAIRE GPT and a DQ Agent to automate rule generation, anomaly detection and issue remediation.

- **Connectivity and partnerships:** Informatica maintains a broad library of metadata-aware connections, enabling support of complex hybrid and multicloud environments. IDMC provides deep integrations across all major hyperscalers and platforms, including AWS, Azure, Google Cloud, Snowflake, Databricks, and Oracle.
- **Global GTM and support:** Informatica's massive global presence across North America, EMEA and APJ are supported by a network of more than 500 global partners, enabling broad implementation and ongoing support options for its customers.

Cautions

- **Uncertainty following Salesforce acquisition:** The Informatica acquisition by Salesforce creates uncertainty regarding Informatica's long-term strategic direction, product offerings, heterogeneous support of non-Salesforce environments, pricing options, and commercial models postacquisition.
- **Learning curve:** Gartner's Peer Insights from Informatica's clients express that there is a steep learning curve to gain proficiency. Remarks suggest that IDMC's flexibility makes the solution challenging to learn how to use all features. Informatica has introduced AI/agentic flows to address this.
- **Cloud-only and end-of-support announcements:** Informatica prioritizes cloud-only solutions (IDMC), raising concerns for some clients about migration challenges and ongoing support for on-premises or hybrid systems. Informatica Data Quality (IDQ) clients should explore support options and/or IDQ to CDQ conversion support via Migration Factory.

Soda

Soda is a Niche Player in this Magic Quadrant, headquartered in Brussels, Belgium. This is its first appearance in the Magic Quadrant. Its data quality products include the commercial Soda Cloud and the open-source Soda Core.

Soda has 150 active commercial customers and over 5,000 active open-source deployments. Its customers are primarily in EMEA and North America, followed by APAC. Its clients are primarily in the financial services, media and retail, and technology sectors. Future focus is on introducing its "Agent Mesh" with specialized AI agents for prioritization, diagnostics and resolution, and a new data management product with a framework designed for operating governed data products at scale.

Strengths

- **Data contracts approach to DQ**: Soda uses "data contracts" to drive all data quality functions. Contracts serve as a formal agreement between those who generate the dataset and those who consume it. A data contract describes the dataset's structure, quality requirements, and SLAs, which standardizes expectations for the dataset's content and quality. Active metadata analysis infers attribute- and cross-attribute validation rules, and AI-assistance automatically generates SQL-based data quality rules from natural language inputs.

- **Proprietary anomaly detection:** Soda's algorithm leverages historical data to dynamically learn and adapt to emerging patterns and anomalies. Peer-reviewed (NeurIPS) accuracy results indicate far fewer false positives over conventional anomaly detection models.

- **Vision pressure tested via open source and cloud hybrid model:** Soda leverages an open-source model for idea generation and validation of new feature concepts, which enables rapid innovation through an agile product development approach.

Cautions

- **Data transformation limitations:** Soda's features for data transformation relies on AI-assisted SQL-based transformations. Out-of-the-box, prebuilt advanced transformation features are not available today. No ability to reference third-party empirical sources or knowledge bases for data standardization or cleansing. Soda does not provide native address validation and enrichment, nor geocode enrichment services. It performs only exact matching or basic matching during downstream data processing. Unstructured data support is limited to metadata collection.

- **Point-of-entry remediation support:** Soda does not support point-of-entry, data quality remediation (standardization, cleansing, and/or matching) functions that are executed

live, within the native business application's environment.

- **Limited conversational AI assistant support:** Soda does not currently offer conversational AI assistants for activities such as interactively requesting the creation of data quality processes and routing assignment to users, or to propose steps to perform data quality tasks or to remediate data quality issues (on the roadmap).

Vendors Added and Dropped

We review and adjust our inclusion criteria for Magic Quadrants as markets change. As a result of these adjustments, the mix of vendors in any Magic Quadrant may change over time. A vendor's appearance in a Magic Quadrant one year and not the next does not necessarily indicate that we have changed our opinion of that vendor. It may be a reflection of a change in the market and, therefore, changed evaluation criteria, or of a change of focus by that vendor.

Added

- **Acceldata:** This vendor is a new entrant to the augmented data quality solutions market. Acceldata has expanded its data quality features, which now meet current inclusion criteria.
- **Soda:** This vendor is a new entrant to the augmented data quality solutions market. This vendor has expanded its data quality features, which meet current inclusion criteria.

Dropped

- **SAS:** The vendor was dropped due to lack of support for unstructured content and for lack of augmentation of critical data quality functions that leverage AI/ML, natural language processing or LLMs in its product offerings at the time of evaluation.

Inclusion and Exclusion Criteria

To qualify for inclusion, vendors must meet all the following inclusion criteria:

- The solution being considered for augmented data quality research must be generally available no later than 24 October 2025.
- Offer stand-alone software solutions that are positioned, marketed and sold specifically for general-purpose data quality applications. Vendors that provide several data quality product components or unified data management platforms must demonstrate that these are integrated and collectively meet the full inclusion criteria for this Magic Quadrant.
- Deliver critical augmented data quality functions at a minimum (descriptions are the same as given in the Market Definition):
  - Profiling, monitoring and detection
  - Rule discovery, creation and management
  - Alerts, notifications and visualizations
  - Data transformations (parsing, cleansing and standardizing data)
  - Matching, linking and merging.
  - Usability, workflow and issue resolution
  - Active metadata and lineage support
  - Unstructured data support
- Support augmentation of the critical data quality functions listed above by leveraging AI/ML features (supervised, semisupervised or unsupervised methods, NLP-based or LLM-supported), graph analysis and metadata analytics (active metadata).
- Support the above functions in both scheduled (batch) and interactive (real-time) modes.
- Enable large-scale deployment via server-based or cloud-based runtime architectures that can support concurrent users and applications. Cloud-based/SaaS version should support all critical data quality functions independently as mentioned in the above criteria.

- Maintain an installed base of at least 50 production paying customers (different companies/organizational entities) for their flagship data quality product (not individual smaller modules or capabilities). The customers must be running in production for at least six months.
- Include a complete solution addressing administration and management, as well as end-user-facing functionality, for four or more of the following types of users: data steward, data architect, data quality analyst, data engineer, database administrator, data integration analyst, data scientist, data analyst, business intelligence analyst and a citizen user.
- Provide out-of-the-box and prebuilt data quality rules based on industry practices for data profiling and monitoring, cleansing, standardization, and transformation.
- Support integrability and interoperability with other data management solutions such as metadata management, master data management, data observability and data integration solutions from third-party tools.
- Provide direct sales and support operations, or a partner providing sales and support operations in at least two of the following regions: North America, South America, EMEA, and Asia/Pacific.
- The customer base for production deployment must include customers in multiple countries and in more than one region (North America, South America, EMEA, and Asia/Pacific), and be representative of at least three or more industry sectors.

The following types of vendors were excluded from this Magic Quadrant, even if their products met the above criteria:

- Vendors that meet the above criteria but are limited to deployments in a single specific application environment, industry or data domain are excluded.
- Vendors who support limited data quality functionalities, no augmentation and automation, or addressing of very specific data quality problems (for example, focusing on only address cleansing and validation) are excluded, because they do not provide the complete suite of data quality functionality expected from today's augmented data quality solutions.

- Vendors who support only on-premises deployment and have no option in cloud-based deployment in any public cloud environment (for example, AWS, Azure or Google Cloud) are excluded.

Evaluation Criteria

Ability to Execute

Gartner analysts evaluate technology vendors on the quality and efficacy of the processes, systems, methods and procedures that enable their performance to be competitive, efficient and effective, and to positively impact their revenue, retention and reputation within Gartner's view of the market.

Gartner evaluates vendors' Ability to Execute in the augmented data quality solutions market by using the following criteria:

- **Product or service:** The capabilities, features and overall quality of the core goods and services that compete in and serve the defined market. These may be offered natively or through OEM agreements/partnerships as defined in the market definition and detailed in the subcriteria. Vendors are evaluated based on their ability to address current market needs with their AI-enhanced data quality capabilities.
- **Overall viability:** The organization's overall financial health, as well as the financial and practical success of the relevant business unit. This includes the likelihood that the organization can continue to offer and invest in the product, as well as the product's position in the organization's portfolio. We consider the vendor's financial strength (based on revenue growth, profitability and cash flow) and the strength and stability of its people and organizational structure. This criterion considers buyers' increased openness to consider newer, less established and smaller vendors with differentiated offerings.
- **Sales execution/pricing:** The organization's capabilities in all presales activities and the structures that support these activities. This includes deal management, pricing and negotiation, presales support, and the overall effectiveness of the sales channel. We consider the effectiveness of a vendor's pricing model in light of current and future customer demands, trends and spending patterns (i.e., operating expenditures and

flexible pricing), as well as the effectiveness of the vendor's direct and indirect sales channels.

- **Market responsiveness/record:** The ability to respond, change direction, be flexible and achieve competitive success as opportunities develop, competitors act, customer needs evolve and market dynamics change. This includes the provider's history of responsiveness to changing market demands. We also consider the degree to which the vendor has demonstrated the ability to respond successfully to market demand for expanded/refined data quality capabilities over an extended period.
- **Marketing execution:** The ability to deliver clear, high-quality, creative and effective messaging via publicity, promotional activity, thought leadership, social media, referrals and sales activities. This includes the organization's ability to influence the market, promote the brand, increase awareness of products and establish a positive reputation among customers. This "mind share" also considers activities driven through partnerships. We consider the overall effectiveness of the vendor's marketing efforts, the level of mind share generated, and the magnitude of market share achieved as a result.
- **Customer experience:** The degree to which a vendor's products, services and programs enable customers to achieve their desired results. This includes the quality of supplier/buyer interactions, technical support or account support, as well as ancillary tools, customer support programs, availability of user groups and service-level agreements. We evaluate the level of satisfaction expressed by the vendor's customers for the vendor's product support, variety of support options, professional services, and the overall relationship with the vendor. We also consider customers' perceptions of the value of the vendor's data quality solution(s) relative to its costs and their expectations.
- **Operations:** The ability of the organization to meet its goals and commitments. This includes the quality of its organizational structure, skills, experiences, programs and systems that enable the organization to operate effectively and efficiently. We consider overall operational health, employee engagement and effective resource management.

Ability to Execute Evaluation Criteria

Evaluation Criteria	*Weighting*
Product or Service	High
Overall Viability	Medium
Sales Execution/Pricing	High
Market Responsiveness/Record	Medium
Marketing Execution	Medium
Customer Experience	High
Operations	Low

Source: Gartner (February 2026)

Completeness of Vision

The evaluation covers current and future market direction, innovation, customer needs and competitive forces, as well as how well they correspond to Gartner's view of the market.

Gartner assesses vendors' Completeness of Vision in the augmented data quality solutions market by using the following criteria:

- **Market understanding:** The ability to understand customer needs and translate that understanding into products and services. Vendors with a clear vision of the market listen to and understand customer demands, and they can shape or enhance market changes with their vision. This includes intuitive support for business-centric roles through advanced/AI-assisted data quality functionality that supports operational and analytic data quality scenarios, invoked through real-time, streaming and batch

processes across structured, semistructured and unstructured content, within hybrid-cloud or multicloud environments. We consider the vendor's vision and execution in support of expanding AI data readiness scenarios. We also consider the interoperability and convergence with other data management-related markets, i.e., data integration, data observability, metadata management and data governance.

- **Marketing strategy:** The ability to clearly communicate differentiated messaging, both internally and externally, through social media, advertising, customer programs and positioning statements. We consider the degree to which a vendor's marketing approach aligns with and/or exploits evolving trends (AI-ready data, business-centric data quality focus and adaptive governance) and the overall direction of the market. We are specifically looking for the volume and content of marketing messages shared via social media channels.

- **Sales strategy:** The ability to create a sound strategy for selling that uses the appropriate networks, including direct and indirect sales, marketing, service and communication. This includes partnerships that extend the scope and depth of a provider's market reach, expertise, technologies, services and their customer base. A vigorous sales strategy includes the alignment of sales models with customer-preferred buying approaches, including freemium programs and subscription-based pricing.

- **Offering (product) strategy:** The ability to approach product development and delivery in a way that meets current and future requirements, with an emphasis on market differentiation, functionality, methodology and features. We consider the degree to which a vendor's product roadmap reflects demand trends, addresses current gaps or weaknesses, and emphasizes competitive differentiation. We also consider the breadth of the vendor's strategy regarding a range of product and service delivery models, from on-premises deployment to data as a service, SaaS, and hybrid and cloud-based models.

- **Business model:** The design, logic and execution of the organization's business proposition. We consider the vendor's overall approach to execute its strategy for the data quality solutions market, including delivery models, funding models (i.e., public or private), development strategy, packaging and pricing options, and partnership types (joint marketing, reselling, OEM, system integration/implementation, etc.).

- **Vertical/industry strategy:** The strategy to direct resources, skills and offerings to meet the specific needs of individual market segments, including vertical markets. The degree of emphasis that the vendor places on vertical-market solutions, and the vendor's depth of vertical-market expertise, including certifications. We specifically consider prebuilt out-of-the-box rules, templates, dashboard or post-implementation training for certain industries.
- **Innovation:** Marshaling of resources, expertise or capital for competitive advantage, investment, consolidation or defense against acquisition. We will examine the vendor's support for or plans to support key and evolving trends, such as GenAI/agentic-AI-driven automation and AI-assistant interactive engagement with users, rules inference and remediation recommendations, intelligence capabilities, deployment flexibility and unstructured content support.
- **Geographic strategy:** The ability to direct resources, skills and offerings to meet the specific needs of regions outside the providers' home region, either directly or through partners, channels and subsidiaries. We consider the strength of a vendor's strategy for expanding into markets beyond its home region or country to support global demand for data quality capabilities and expertise.

Completeness of Vision Evaluation Criteria

Evaluation Criteria	*Weighting*
Market Understanding	High
Marketing Strategy	Medium
Sales Strategy	Medium
Offering (Product) Strategy	High
Business Model	Low

Evaluation Criteria	*Weighting*
Vertical/Industry Strategy	Medium
Innovation	High
Geographic Strategy	Low

Source: Gartner (February 2026)

Quadrant Descriptions

Leaders

Leaders demonstrate substantial depth across the full range of data quality functions, including the core data quality capabilities that have existed for years, as well as bringing more automation and AI augmentation to data quality solutions.

Leaders exhibit a clear understanding of dynamic trends in the data quality market. They explore and execute thought-leading and differentiating ideas, and they deliver product innovation based on market's demands. Leaders also provide capabilities for deeper data insights by leveraging advanced technologies such as AI/ML/LLMs, AI agents/agentic AI, knowledge graph, active metadata or NLP to minimize manual effort.

Leaders align their product strategies with the latest market trends. These trends include focusing on a nontechnical audience, trust-based governance, growth in data diversity, low data latency, in-depth data quality analytics (not just reporting) and intelligent capabilities. Other trends are new delivery options (such as cloud, hybrid cloud and IoT edge deployment), and alternative pricing and licensing models (such as meter-based, consumption-based or pay-as-you-go, or persona/use-case-based).

Leaders address all industries, geographies, data domains and use cases. Their products support multidomain and alternative deployment options such as SaaS or microservices. They offer excellent support for business roles with easy-to-use visualization, and they include out-of-the-box, AI-augmented and machine learning capabilities and predictive analytics.

Leaders offer extensive support for a variety of traditional and new data sources and formats (including cloud platforms, data lake, IoT platforms, open table format), a trust-based governance model, and delivery of enterprise-level data quality implementations.

Leaders have significant organizational size, an established market presence and a multinational presence (either directly or through a parent company). Leaders also undertake clear, creative and effective marketing, which influences the market, promotes their brand and increases their mind share.

Challengers

Challengers exhibit a strong understanding of the current demands of the augmented data quality market as well as both the credibility and viability to deliver. They have solid sales and marketing execution. Therefore, they generally have substantial customer bases. Some Challengers are mature in specific capabilities and use cases, which enables them to deliver targeted use cases faster and with a better overall total cost of ownership than other vendors (sometimes even Leaders). Some may also focus on certain ecosystems. These vendors have developed best practices for leveraging their strongest product capability in new delivery models.

Challengers may not have the same breadth of offering as Leaders and/or, in some areas, may not demonstrate as much thought leadership and innovation. For example, they may focus on a limited number of data domains (e.g., customer, product or location data). They may not exhibit full understanding of concepts like augmented data quality and may not have features related to these as part of their vision. They may not have built-in metadata management capabilities but can provide integration with metadata management platforms. The data observability features are foundational, but not comprehensive.

Challengers may lack capabilities in areas such as real-time profiling on streaming data, AI-assisted anomaly detection, predictive analysis or support of complicated data landscapes. Leveraging LLMs and NLP is rudimentary or progressing slowly.

Compared with Leaders, Challengers often exhibit less understanding of some areas of the market, and their product strategies may suffer from a lack of differentiation.

Visionaries

Visionaries are innovators and demonstrate a strong understanding of emerging technology and business trends, or they focus on a specific market need that is outside of common practices, while also offering capabilities that are expected to grow in demand. They are aligned with the market in adding features related to automation and augmentation as a part of their roadmaps. They have a vision toward more unified platforms that allow convergence of capabilities or integration plans with adjacent markets for automation of data quality processes.

Visionaries should lead the push toward the utilization of knowledge graphs, semantics, active metadata and AI/ML/LLM/NLP for significant automation in data quality design, remediation delivery and monitoring. Visionaries also focus on a nontechnical audience, trust-based governance, growth in data diversity, low data latency, data quality analytics and intelligent capabilities. They include new delivery options (such as container-based or IoT edge deployment) and alternative pricing models (such as open source and subscriptions). Visionaries' product capabilities are mostly aligned with these trends, but not as completely as Leaders.

Although Visionaries can deliver good customer experiences, they may lack the scale, market presence, brand recognition, customer base and resources of Leaders. They have good vision but are relatively slow in execution of these great vision elements.

Niche Players

Niche Players often specialize in a limited number of industries, geographic areas, market segments (such as small and midsize businesses) or data domains (such as customer data or product data). They often have strong offerings for their chosen areas of focus and deliver substantial value for customers in those areas. Niche Players may not appear frequently in competitive situations for comprehensive and/or enterprise-class data quality deployments. Many have strong offerings for a specific range of data quality challenges.

Niche Players typically have limited market share and presence, and have limited functionalities or lack financial strength. Niche Players often have to catch up with the latest innovations, such as active metadata support, machine learning and observability.

Niche Players often exhibit advantages in pricing within their established footprint and in vertical or horizontal solutions, but sometimes cannot complement an organization's other data management technologies. Niche Players are known for solving one part of the data quality problems well through a targeted solution and may be the right solution for organizations with less complex needs.

Context

The acceleration of AI, GenAI, and now AI agent adoption and the dynamics of business change have escalated the demand for greater scale and shorter time-to-value in relation to data management. Data quality is a particular concern, because trusted, high-quality data is vital to the success of growing AI and business initiatives. According to the 2026 Gartner CIO and Technology Executive Survey, 81% of respondents reported their enterprise would increase funding for traditional AI and 84% would increase funding for generative AI in 2026. [1] D&A leaders and their teams are responsible for delivering foundational AI-ready data and governance.

To put that into perspective, according to the 2024 Gartner AI Mandates for the Enterprise Survey, on average, about 40% of AI prototypes make it into production, and participants reported data availability and quality as a top barrier to AI adoption. [2] In addition, more than 64% of data management leaders stated that data quality and data governance remain one of their top five investment areas in the next two to three years, according to Gartner's 2024 Gartner Evolution of Data Management Survey. [3]

All these challenges drive the adoption of augmented data quality solutions. Data is useful only if its quality, content and structure are documented and well-understood. The cost of dirty, insufficient and/or inaccurate data remains a substantial threat. Delivering reliable, trusted and timely data for business consumption and for AI model training and testing is a continuous effort and process that can be supported with modern technologies in augmented data quality solutions.

In this report, we assess 13 vendors. Some are more advanced in augmentation and automation, and some are picking up speed toward the same goal. Use this Magic Quadrant to help find the right vendor and product for your organization's needs. Gartner strongly

advises against selecting a vendor solely because it is in the Leaders quadrant. A Challenger, Niche Player or Visionary could be the best match for your requirements. Use this Magic Quadrant in combination with the companion **Critical Capabilities for Augmented Data Quality Solutions**, as well as Gartner's client inquiry service and Peer Insights portal.

Market Overview

The market for augmented data quality solutions continues to experience consistent expansion, maintaining its dynamism as requirements evolve and intensify, particularly with the rise of AI, GenAI, AI agents, and the imminent adoption of agentic AI. Ongoing digital transformation efforts also fuel this growth. The introduction of AI-driven technologies is reshaping the augmented data quality landscape, leading to new approaches in managing data quality processes and extracting insights.

Augmented data quality solutions, powered by AI, GenAI, AI agents and metadata, automate data quality tasks to accelerate value realization and extend the depth of data comprehension. Top vendors are enhancing their platforms with advanced automation and deeper insights. Two major trends remain central: the use of AI supports data quality work and the necessity of data quality to support AI work.

AI Supports Data Quality Work

AI technologies are transforming the data quality life cycle — spanning discovery, assessment, association, validation, correction (cleansing, standardization, matching and merging), and monitoring — by enabling more efficient management of individual datasets. Machine learning algorithms analyze data to identify anomalies and infer rules through functional dependency analysis, while clustering techniques automate dataset cleansing by detecting outliers and recommending corrections. This methodology facilitates faster and more thorough data quality interventions at the dataset level.

Vendors are deploying a blend of supervised, semisupervised and unsupervised learning models to bolster data quality capabilities and streamline operations. Most mainstream solutions use supervised learning where data entities and relationships are well-defined,

while unsupervised models autonomously detect patterns and outliers. For instance, unsupervised matching algorithms are used to match customer records by learning from data attributes and user feedback.

Natural language processing and large language models excel at interpreting, parsing and managing human language. Within data quality, NLP is instrumental in profiling, parsing, matching, standardizing and cleansing data using natural language inputs. Business users can describe new data quality requirements conversationally, such as stating, “Product A’s height must equal 10 inches.”

Vendors in the data quality space are integrating GenAI capabilities, providing ChatGPT-like features with their solutions. Vendors prioritize security requirements with many using commercial OpenAI platforms (e.g., Microsoft Azure OpenAI Service, Google Vertex AI). Vendors also offer proprietary interfaces for LLM input/output translation.

AI agents are being incorporated to further automate data quality processes. Beyond identifying issues, these agents can semiautonomously (human-in-the-loop) or fully autonomously correct and resolve data quality problems. Multiagent orchestration is emerging, with specialized agents dedicated to tasks like root cause analysis, rule discovery and impact-based alerting. A point of awareness regarding AI agents and agentic AI: these technologies are very recent additions to the technology stack. Therefore, the hype associated with AI-assisted solution capabilities should be carefully evaluated for tangible results based on your business scenarios and data quality requirements.

A recent advancement is the implementation of the Model Context Protocol, an open standard facilitating interoperability and context sharing among AI agents across hybrid environments and data platforms. Some vendors now offer MCP servers, enabling users to access data quality metadata through external clients such as Claude or Microsoft Copilot.

For example, a business user might query an LLM about a dataset’s health, receiving quality scores and lineage information from the data quality vendor’s platform via MCP. Select vendors use MCP to enable internal AI agents to interact with agents from other solutions, exchanging data quality insights and triggering related actions. The future vision is a fully autonomous, multiagent ecosystem capable of self-healing data environments.

Data Quality Supports AI Work

Reliable and high-quality data are fundamental for developing robust AI systems. Augmented data quality solutions safeguard data integrity for all types of AI applications, including emerging agentic AI and guardian agents, by offering profiling, monitoring and detection capabilities throughout the data pipeline. Integration with external data sources also enables data enrichment, enhancing overall data quality.

Strong attention is now given to the quality and context of unstructured data content, including the evaluation of sensitive or personal content and the identification of content uniqueness. This strengthens connections within data pipelines, preparing them for retrieval augmented generation (RAG), fine-tuning, and providing additional business context via MCP.

To address heightened requirements for data sovereignty driven by global regulations, cybersecurity concerns, geopolitical factors and industry-specific compliance, some vendors now offer virtual private cloud (VPC) deployment options.

Additional Use Cases Demand Data Quality

- **AI-ready data:** Augmented data quality platforms deliver technologies to prepare data for AI applications, including assessment, schema and quality monitoring, accuracy validation, and error correction or dataset preparation for AI (see **A Journey Guide to Deliver AI Success Through AI-Ready Data**).
- **Data contracts:** These solutions are evolving to enforce data quality via data contracts, allowing business users to define expectations in natural language, which are then translated into executable validation rules at ingestion. Pipelines can automatically reject data failing to meet these standards.
- **Data products:** Augmented data quality ensures that datasets used for data products are accurate, reusable, shareable and compliant with relevant regulations or policies.

Unstructured/Semistructured Content Support

With the proliferation of AI, innovation around unstructured content is advancing rapidly. Data and analytics leaders are increasingly leveraging unstructured data for RAG, with data

quality a critical component (see **Governing Unstructured Data for AI Readiness: A Strategic Roadmap**). Unstructured content also enriches data quality actions through MCP.

Augmented data quality solutions now analyze unstructured and semistructured data, identifying quality issues via semantic analysis and business validation logic, and generating context-aware metadata. AI, ML, LLM, NLP and graph technologies are utilized to evaluate the accuracy, completeness and consistency of unstructured data, based on metadata availability and validation logic. Vendors typically pursue one or more of the following strategies:

- Co-development with hyperscalers
- Adoption of open-source LLMs
- Customization or fine-tuning of commercial LLMs
- Development of proprietary LLMs

Market Performance and Growth

The data quality solutions sector has demonstrated robust revenue growth, reaching $2.2 billion in 2024 (see **Market Share Analysis: Data Management Software (Excluding DBMS), Worldwide, 2024**). Augmented data quality solutions accounted for 16% of the total data management solutions market share in 2024.

In 2024, the leading three vendors — SAP, Experian and Precisely — held a combined 46.7% market share, a figure largely unchanged from 2023 (see **Market Share: Data and Analytics Software, Worldwide, 2024**). This concentration, consistently around 45% over the past four years, highlights the dominance of top vendors in the market.

The market is divided into traditional and augmented data quality solutions. SAP, as a traditional vendor, leads in market share due to its extensive customer base. However, vendors with a strong vision and roadmap for augmented capabilities are now favored, while those lacking such direction are considered to be falling behind. Competitive advantage increasingly hinges on investment in augmented data quality, with laggards at risk of obsolescence as the market consolidates.

Market Dynamics: Convergence

Integration With Data Governance and Data Management Tools

Data quality is increasingly merging with other segments such as data management and governance. Many augmented data quality solutions tracked by Gartner also offer products in D&A governance, metadata management, master data management, and data integration (see **Magic Quadrant for Data and Analytics Governance Platforms**, **Magic Quadrant for Metadata Management Solutions**, **Market Guide for Master Data Management Solutions** and **Magic Quadrant for Data Integration Tools**). This convergence reflects the holistic demand for data quality throughout the data life cycle.

Integration With Data Observability Tools

The emergence of data observability introduces a new technological dimension to augmented data quality. Data observability enables organizations to assess the overall health of their data environments (see **Market Guide for Data Observability Tools**). This trend extends augmented data management by combining features from augmented data quality, active metadata and DataOps.

Data observability solutions focus on automated anomaly and outlier detection with algorithms that can be repurposed for rule creation. Many data quality vendors have integrated advanced observability features, though stand-alone observability tools typically lack remediation capabilities. As a result, partnerships between data quality and observability providers create a comprehensive ecosystem for end-to-end data quality management.

Acronym Key and Glossary Terms

Trust-based governance	This governance approach leverages a structured framework that assesses, communicates, and manages the trustworthiness of data assets using graduated levels. It integrates elements of data quality — including dimensions such as completeness, consistency and accuracy — with an additional “believability” or trust dimension. Trust-based governance

	includes several key tenets: the measurement and grading of trust in data assets, the integration of trust metrics with formal decision rights models, and the requisite investment in technology that supports dynamic, scalable governance processes.

⊕ Evidence

⊕ Evaluation Criteria Definitions

Unlock more AI value

With unmatched, proprietary AI insights from Gartner

Explore the AI Hub ↗

© 2026 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by Gartner's Usage Policy. Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "Guiding Principles on Independence and Objectivity." Gartner research may not be used as input into or for the training or development of generative artificial intelligence, machine learning, algorithms, software, or related technologies.

© 2026 Gartner, Inc. and/or its Affiliates. All Rights Reserved.